Generalized Shrinkage and Penalty Functions

Rick Chartrand Los Alamos National Laboratory rickc@lanl.gov

Abstract—We extend the proximal mapping property of soft thresholding to a general class of shrinkage mappings. We give an example and demonstrate improved reconstruction performance.

I. INTRODUCTION

A key ingredient in many algorithms for ℓ^1 minimization is that the ℓ^1 norm has an explicit, simple proximal mapping:

$$\underset{w}{\arg\min} \|w\|_1 + \frac{1}{2\lambda} \|w - x\|_2^2 = \max\{|x| - \lambda, 0\} \operatorname{sign}(x). \tag{1}$$

However, there is much research showing that better results for sparse signal recovery can be obtained using nonconvex penalty functions, such as the ℓ^p quasinorm for $p \in (0,1)$ —which does not have an explicit proximal mapping for general p. The ℓ^0 penalty does have an explicit proximal mapping, though it is discontinuous, and difficult to use numerically.

To bridge the gap between ℓ^p quasinorms and proximal mappings, in [1] a new family of shrinkage operators was proposed (cf. [2]), generalizing (1):

$$S_p(x,\lambda) = \max\{|x| - \lambda^{2-p}|x|^{p-1}, 0\} \operatorname{sign}(x).$$
 (2)

It was then shown [3] that for $p \in (-\infty, 1]$, there is a penalty function G_p having S_p for a proximal mapping, and that G_p has many desirable properties, with better numerical results being obtained with p < 1.

In this paper, we generalize this process, by giving conditions under which a mapping will be the proximal mapping of a penalty function that will be useful for sparse recovery. We then give an example of such a mapping, and demonstrate improved image reconstruction using it.

II. SHRINKAGES AND PENALTY FUNCTIONS

We generalize the notion of a shrinkage mapping:

Theorem 1. Suppose $s:[0,\infty)\to\mathbb{R}$ is continuous, satisfies $x\leq\mu\Rightarrow s(x)=0$ for some $\mu\geq0$, is strictly increasing on $[\mu,\infty)$, and $s(x)\leq x$. Define S on \mathbb{R}^n by $S(x)_i=s(|x_i|)\operatorname{sign}(x_i)$ for each i. Then S is the proximal mapping of a penalty function $G(x)=\sum_i g(x_i)$ where g is even, strictly increasing and continuous on $[0,\infty)$, differentiable on $(0,\infty)$, and nondifferentiable at 0 iff $\mu>0$. If also x-s(x) is nonincreasing on $[\mu,\infty)$, then g is concave on $[0,\infty)$ and G satisfies the triangle inequality.

Proof: Define $f(x)=\int_0^{|x|}s(t)\,dt$. Then f is C^1 and even, and since $f'(x)=s(|x|)\,{\rm sign}(x),\ f$ is convex, and

This research was supported by the U.S. Department of Energy through the LANL/LDRD Program.

strictly convex on $(-\infty, -\mu]$ and $[\mu, \infty)$. Define g by $g(w) = (f^*(w) - w^2/2)/\lambda$, where f^* is the Legendre-Fenchel transform of f [4, p. 473]. Then by [4, Prop. 11.3], we have $f'(x) = \arg\max_w wx - f^*(w)$. Then $s(x) = f'(x) = \arg\min_w g(w) + \frac{1}{2\lambda}(w-x)^2$ by simple manipulations. Thus $S(x) = \arg\min_w G(w) + \frac{1}{2\lambda}\|w-x\|_2^2$ by separability, showing that S is the proximal mapping of G.

The hypotheses on s allow the properties of g to follow as in [3, Prop. 3], *mutatis mutandis*, except here we obtain $\partial g(0) = [-\mu/\lambda, \mu/\lambda]$, and the nondifferentiability claim follows.

III. EXAMPLE

We consider $s(x) = x \exp(-\alpha/(\exp(x-\lambda)-1)^2)$ for $x > \lambda$ and s(x) = 0 on $[0, \lambda]$, a smooth approximation of hard thresholding. We apply Thm. 1 and solve $\min_x G(\nabla x)$ subject to $(\mathcal{F}x)|_{\Omega} = (\mathcal{F}s)|_{\Omega}$, where \mathcal{F} is the 2D DFT, the sampling set Ω consists of 6 radial lines, and s is the 256×256 Shepp-Logan phantom. We use a split Bregman algorithm as in [1], [5], but with the new shrinkage substituted (with $\alpha = 10^{-2}$, $\lambda = \mu = 10^{-10}$). The result (Fig. 1) is a perfect reconstruction, from fewer data than the previous fewest [6].

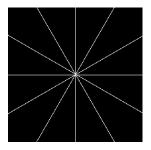




Fig. 1. Left: sampling mask (courtesy W. Guo [6]), with 2.56% sampling. Right: reconstruction using the new shrinkage function is exact, from fewer data than ever before.

REFERENCES

- [1] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *IEEE International Symposium on Biomedical Imaging*, 2009.
- [2] R. Chartrand and B. Wohlberg, "A nonconvex ADMM algorithm for group sparsity with sparse groups," in *IEEE International Conference* on Acoustics, Speech, and Signal Processing, 2013.
- [3] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, pp. 5810–5819, 2012
- [4] R. T. Rockafellar and R. J.-B. Wets, Variational Analysis. Berlin: Springer-Verlag, 1998.
- [5] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," SIAM J. Imaging Sci., vol. 2, pp. 323–343, 2009.
- [6] W. Guo and W. Yin, "Edge guided reconstruction for compressive imaging," SIAM J. Imaging Sci., vol. 5, pp. 809–834, 2012.